DOCUMENT RESUME

ED 078 067 TM 002 886

AUTHOR	Whitlock, Gerald H.; Cureton, Edward E.
TITLE	Validation of Morale and Attitude Scales. Technical Report.
INSTITUTION	Air Research and Development Command, Lackland AFB, Tex. Wright Air Development Div.
REPORT NO	WADD-TR-60-76
PUB DATE	Jun 60
NOTE	31p.; Report prepared under contract with the University of Tennessee, Knoxville
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Attitudes; Comparative Analysis; Interviews; *Military Personnel; Prediction; Questionnaires; *Rating Scales; Technical Reports; Test Construction; *Test Validity

ABSTRACT

This paper completes reporting of the develcpment of scales for measuring morale among Air Force personnel. A previously developed questionnaire keyed for eight scales was given tc an independent sample. When these scores were compared with criteria derived from interviews, ratings, and Air Force records, validities were uniformly low. Although none of the validities are high enough for useful prediction, the scales do measure expressed attitudes with considerable reliability. (Author)

FILMED FROM BEST AVAILABLE COPY

WADD-TR-60-76

June 1960

Technical Report

ED 0,8067

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Validation of Morale and Attitude Scales

TM 002 886

By
Gerald H. Whitlock
Edward E. Cureton

University of Tennessee
Contract AF 41(657)-247

PERSONNEL LABORATORY
WRIGHT AIR DEVELOPMENT DIVISION
AIR RESEARCH and DEVELOPMENT COMMAND
UNITED STATES AIR FORCE
LACKLAND AIR FORCE BASE, TEXAS

Fred E. Holdrege, Col USAF
Chief

A. Carp
Technical Director

Personnel Laboratory
Wright Air Development Division (ARDC)

NOTICES

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely related Government procurement operation, the United States Government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

Qualified requestors may obtain copies of this report from the Armed Services Technical Information Agency, Documents Service Center, Arlington 12, Virginia. Department of Defense contractors may obtain unclassified documents from ASTIA on request by stating their official need and citing the Defense contract involved.

This report has been released to the Office of Technical Services, U.S. Department of Commerce, Washington 25, D.C., for sale to the general public.

Copies of WADD Technical Reports and Technical Notes should not be returned to the Wright Air Development Division unless such return is required by security considerations, contractual obligations, or notice on a specific document.

ED 078067

WADD-TR-60-76
June 1960

VALIDATION OF MORALE AND ATTITUDE SCALES

By
Gerald H. Whitlock
and Edward E. Cureton

University of Tennessee

Project 7719, Task 17130
Contract AF 41(657)-247

Personnel Laboratory
WRIGHT AIR DEVELOPMENT DIVISION
AIR RESEARCH AND DEVELOPMENT COMMAND
UNITED STATES AIR FORCE
Lackland Air Force Base, Texas

Foreword

The project for developing techniques of measuring morale among Air Force personnel was carried out under Contract AF 41(657)-247 with the University of Tennessee, Knoxville. The Principal Investigator was Dr. Edward E. Cureton, head of the Department of Philosophy and Psychology, University of Tennessee. The Project Director was Dr. Gerald H. Whitlock, associate professor, University of Tennessee. The Contract Monitor was Dr. Lois L. Elliott, now with the School of Aviation Medicine, Brooks Air Force Base.

Other reports issued under this contract are:

The Status of Morale Measurement, 1959, by Gerald H. Whitlock.
(WADD-TN-60-136, May 1960)

Dimensions of Airman Morale, by Edward E. Cureton.
(WADD-TN-60-137, June 1960)

Factor-analytic Reanalysis of Studies of Job Satisfaction and Morale, by Edward E. Cureton & Bryan B. Sargent.
(WADD-TN-60-138, July 1960)

Mr. John Motley, Air Force Base Educational Officer, was responsible for the smooth and effective scheduling and coordination of test administration and interviews which provided the data for this paper.

Special acknowledgment is due to Mr. Julian T. Allison, Research Assistant, who performed all of the sorting, tabulating, and computing operations for this report; to Mr. Robert S. Hoeke, Research Assistant, who did about half the field work; and to Mrs. Evelyn C. Ottaway, Secretary, who managed the business aspects of the project and typed all correspondence and reports.

Other members of the project staff were:

James M. Porter, Jr., Industrial Psychologist
John B. Knox, Industrial Sociologist
Raymond R. Shrader, Social Psychologist
Louis E. Dotson, Sociometrist
Rolf E. Bargmann (Virginia Polytechnic Institute), Statistician
Albert M. Craig, Research Assistant
Bryan B. Sargent, Research Assistant
Robert F. Shanks, Research Assistant
Michael Deleanu, Research Assistant
Harold E. Edwards, Research Assistant
Harry J. Bruce, Research Assistant
Monroe P. Friedman, Research Assistant
Robert S. Tacker, Research Assistant

Abstract

This paper completes reporting of the development of scales for measuring morale among Air Force personnel. A previously developed questionnaire keyed for eight scales was given to an independent sample. When these scores were compared with criteria derived from interviews, ratings, and Air Force records, validities were uniformly low. Although none of the validities are high enough for useful prediction, the scales do measure expressed attitudes with considerable reliability.

Table of Contents

	Page
The Revised Questionnaire	1
Construction	1
Keying	1
The Sample	2
The Scales	2
Frequency Distributions	2
Basic Statistical Data	5
Criterion Data	6
Data from Official Records	6
Data from Personal Interviews	6
Quantification of Interview Data	7
Validity Analyses	7
Citations	8
Airman Performance Reports	8
Proficiency Pay	9
Injuries	9
Dispensary Visits	9
Disciplinary Actions	9
Job Performance (Supervisors' Ratings)	10
Rated Morale (Supervisors' Ratings)	10
Rated Group Morale (Supervisors' Ratings)	10
Absences (Supervisors' Estimates)	11
Compatibility of Work with Training	11
Performance Rankings	11
Military Rank	12
Size of Work Group	12
Race	12
Marital Status	13
Number of Dependents	13
Age and Length of Service	13
Education	14
Aptitude Test Scores	14
Airman Proficiency Test	14
Assignment and Career Field	15
Length of Time in Squadron	15
Overall Value to the Air Force	15

Table of Contents (*Continued*)

	Page
Summary	16
Interpretation and Conclusions	16
Scaling	16
Validity	18
References	19
Appendix: Item Numbers Identifying the Same Item in Form A and Revised Questionnaires	20

Validation of Morale and Attitude Scales*

This is the last of four papers reporting outcomes of an investigation aimed at developing effective means for measuring morale among Air Force personnel. The first report of this series (Whitlock, 1960) reviewed the literature on the definition and measurement of morale and job satisfaction, with particular reference to theoretical considerations and to factor-analytic studies of job attitudes and morale. The second report (Cureton, 1960) described the construction of a 167-item questionnaire designed to measure all possible aspects of airman morale, and presented cluster-analysis and factor-analysis data from its administration to 1000 airmen at four Air Force bases. A third paper (Cureton & Sargent, 1960) reported the factoring and refactoring of a number of previous studies giving intercorrelations among job-attitude and morale scales, attempted to synthesize the results of these studies, and included notes on the special precautions required when small correlation matrices are factor-analyzed.

On the basis of the data from the second report, a revised questionnaire was prepared. The present report gives the results of the administration of this revised questionnaire to a new sample of approximately 500 airmen.

The Revised Questionnaire

Construction

The questionnaire used in the first tryout contained 167 questions. The second study of this series reported the development of 17 initial clusters. These clusters included 136 of the original 167 items. All items for which the factor analysis yielded communalities less than .10 were omitted from the clusters. They either have low reliabilities or measure attitudes so specific that they are of little value as scale items. A number of additional items were omitted because some of the clusters were large enough without them.

The revised questionnaire consisted of these 136 items. They were rearranged into a new order such that no two from the same cluster were close together, and those from each cluster were fairly well distributed throughout the questionnaire.

The Appendix relates the item numbers of Form A of the original questionnaire to the item numbers of the revised questionnaire. A copy of Form A is appended to the second report (Cureton, 1960).

Keying

After the revised questionnaire was completed, it was found that the 17 original clusters were highly overlapping, and that several of them measured approximately what was measured by the original (167-item) questionnaire as a whole.

From the overlap data and the data from a preliminary factor analysis of the 17 clusters, eight new cluster-scales were derived:

1. Satisfaction with the Supervisor: 12 items.
2. Satisfaction with the Air Force as a Military Organization: 7 items.
3. Job Satisfaction: 8 items.
4. Satisfaction with the Civilian Community and with the Attitudes of Civilians toward Airmen: 5 items.

*Released by the authors for publication as a WADD Technical Report in March 1960.

5. Satisfaction with the Air Force as a Whole (i.e., with Air Force life in general): 12 items.
6. Satisfaction with Management and Communication (mainly the latter): 12 items.
7. Satisfaction with the Unit and its Leadership (co ..orkers and squadron officers and non-coms): 11 items.
8. General Morale: 25 items measuring whatever was measured, on the average, by the original 167-item questionnaire as a whole.

Scales 1-7 inclusive should be relatively independent of one another; i.e., their intercorrelations should be substantially lower than their reliabilities, though they should all be positive and considerably above zero. Scales 1-4 inclusive should be relatively independent of Scale 8; Scales 5-7 inclusive should be too highly correlated with Scale 8 to be useful in a battery which includes Scale 8. Scale 2 should have relatively low reliability; the other seven should have satisfactory reliabilities. These statements are based on the data from the previous study, and subject to verification or refutation by the data of the present study.

Scales 1-7 inclusive contain no common items. Scale 8 contains four items from Scale 5, four from Scale 6, and two from Scale 7. Of the 136 items in the revised questionnaire, only 82 are scored, and ten of these are scored on both Scale 8 and one other scale.

The Sample

The revised questionnaire was administered to 555 airmen. Of these, 44 were eliminated from further study because of inability to locate personnel records, and the questionnaires of 31 were not available in time for the main analysis, leaving a total of 480 cases for the present analysis.

The sample was selected by requesting from each squadron at one Air Force base a number of men proportional to the size of the squadron, and the squadrons were permitted to select the airmen who were to participate. Airmen of the rank of master sergeant and above were excluded.

The questionnaire was administered to groups of approximately twenty, with no group exceeding thirty. Prior to each session a standard explanation of the project was given, after which the examiner read aloud the instructions on the cover page of the morale questionnaire. The average time required was slightly under one hour. Administration began August 13, and the last stragglers were given the questionnaire on August 21, 1959.

The Scales

Frequency Distributions

Table 1 gives the raw-score frequency distributions for the eight scales. For each item, responses could range from 1 (favorable attitude or high morale) to 5 (unfavorable attitude or low morale). The lowest possible score for each scale is the number of items in the scale, and the highest possible score is five times the number of items. Note that *low* scores indicate favorable attitudes or high morale.

TABLE 1. Frequency Distributions of Raw Scores on the Eight Scales

Scale 1		Scale 2		Scale 3		Scale 4	
Score	f	Score	f	Score	f	Score	f
54-55	1	26	2	40-41	1	25	4
52-53	2	25	1	38-39	9	24	2
50-51	4	24	0	36-37	11	23	6
48-49	11	23	2	34-35	14	22	3
46-47	5	22	4	32-33	13	21	14
44-45	9	21	4	30-31	28	20	17
42-43	15	20	2	28-29	35	19	23
40-41	20	19	8	26-27	39	18	27
38-39	21	18	4	24-25	45	17	34
36-37	25	17	15	22-23	48	16	35
34-35	21	16	19	20-21	59	15	49
32-33	35	15	29	18-19	58	14	41
30-31	33	14	37	16-17	39	13	54
28-29	42	13	52	14-15	34	12	46
26-27	39	12	50	12-13	31	11	31
24-25	31	11	64	10-11	15	10	40
22-23	40	10	73	8- 9	1	9	20
20-21	41	9	54		480	8	15
18-19	30	8	45			7	11
16-17	24	7	15			6	8
14-15	22		480				480
12-13	9						
	480						

Scale 5		Scale 6		Scale 7		Scale 8	
Score	f	Score	f	Score	f	Score	f
59-61	1	51-52	2	53-54	1	108-111	2
56-58	7	49-50	2	51-52	0	104-107	3
53-55	8	47-48	5	49-50	2	100-103	2
50-52	20	45-46	8	47-48	1	96- 99	7
47-49	25	43-44	13	45-46	6	92- 95	13
44-46	27	41-42	11	43-44	9	88- 91	16
41-43	40	39-40	27	41-42	14	84- 87	26
38-40	42	37-38	38	39-40	25	80- 83	41
35-37	64	35-36	43	37-38	36	76- 79	35
32-34	62	33-34	45	35-36	37	72- 75	47
29-31	53	31-32	59	33-34	35	68- 71	43
26-28	47	29-30	34	31-32	42	64- 67	47
23-25	35	27-28	35	29-30	38	60- 63	44
20-22	29	25-26	48	27-28	47	56- 59	40
17-19	10	23-24	34	25-26	38	52- 55	27
14-16	6	21-22	21	23-24	41	48- 51	32
11-13	4	19-20	23	21-22	32	44- 47	27
	480	17-18	14	19-20	37	40- 43	15
		15-16	13	17-18	16	36- 39	5
		13-14	5	15-16	15	32- 35	5
			480	13-14	8	28- 31	3
					480		480

TABLE 2. Basic Statistical Data for the Eight Scales

(N = 480)

Scale		1	2	3	4	5	6	7	8
	r		.10	.35	.23	.45	.52	.61	.59
1	r_c		.13	.39	.29	.50	.60	.69	.64
	Cos		.45	.46	.39	.50	.60	.68	.73
	r			.14	.11	.15	.14	.15	.17
2	r_c			.18	.16	.20	.18	.20	.23
	Cos			.68	.49	.66	.67	.60	.75
	r				.28	.53	.41	.58	.58
3	r_c				.36	.60	.48	.67	.66
	Cos				.43	.67	.53	.64	.71
	r					.42	.39	.38	.46
4	r_c					.53	.51	.49	.57
	Cos					.69	.73	.71	.75
	r						.62	.58	.85
5	r_c						.71	.66	.93
	Cos						.69	.68	.90
	r							.65	.83
6	r_c							.76	.94
	Cos							.76	.90
	r								.76
7	r_c								.86
	Cos								.87
3n		36	21	24	15	36	36	33	75
M		28.50	11.95	22.38	14.15	34.62	30.34	29.12	67.01
σ		9.26	3.33	6.79	7.84	9.53	7.74	7.77	15.43
KR20		.92	.65	.85	.70	.91	.84	.86	.92
Rep		.57	.56	.50	.49	.51	.46	.48	.44

All of the distributions are skewed toward the top, with the higher frequencies associated with the lower scores. This indicates that for this sample the attitudes measured are in general favorable, and the morale high. The skewness is particularly marked in the case of Scale 2 (Military), indicating that these men regard the mission of the Air Force as important, its capabilities high, and their own contributions to the Air Force as a military organization substantial.

Basic Statistical Data

Table 2 gives the basic statistical data for the eight scales. The first complete row at the bottom contains entries 3n: three times the number of items in the scale. If each distribution were symmetric about the score half-way between the lowest possible score and the highest possible score, the 3n value would be the mean.

The next row gives the actual means. All of them are lower than the 3n values, indicating again that the mean attitudes tend in general to be favorable. The largest discrepancy is for Scale 2 (Military); this confirms the previous observation concerning the skewness of this scale.

The following row contains the standard deviations. Here again the value for Scale 2 is conspicuously low, compared particularly with the value for Scale 4. Scale 2 has seven items; Scale 4 has only five.

The next row contains the reliability coefficients, computed by the generalized Kuder-Richardson Formula 20 (generalized because the item-variances do not reduce to values of pq as they do for items scored 0 or 1). The reliability of Scale 2 is relatively low, as would be expected in view of its low standard deviation. Those of Scales 1 (Supervision) and 5 (Air Force as a Whole) are comparatively high for 12-item scales. That of Scale 8 (General Morale), with 25 items, is no higher.

The last row contains the coefficients of reproducibility. In computing them we departed somewhat from Guttman's procedure. For each item, the five sets of response frequencies were distributed according to total scores. Response categories were *not* combined, and every cutting point was required to fall between two numerically different total scores. Guttman's requirement that the total error must not exceed the total non-error in any category of any item was followed rigidly; in several cases this increased the total error over what it would have been had the cut been so placed as to equalize the "above" error and the "below" error. The reported reproducibility of each scale is the mean reproducibility of its items. None of the scales even approach Guttman's criterion of homogeneity: .85 to .90. These scales are at best quasi-scales in his terminology, if they are even that. The rank correlation between the eight reliability coefficients and the eight corresponding coefficients of reproducibility is −.05.

In the main body of the table, the first entry of each set of three is the raw correlation (product-moment) between the two scales. The second is the correlation corrected for attenuation. Since the items of each scale were distributed throughout the questionnaire, the scales were in effect administered simultaneously, and KR20 is an appropriate reliability coefficient to use in making the correction. The last entry of each set is the cosine of the angle between the cluster-centroid vectors of the two scales, based on the data of the previous study (N = 1000), except that in column 8 it is the cosine of the angle between the cluster-centroid vector and the first centroid axis. This is almost, but not quite, the same as the cosine with the centroid of the cluster represented by Scale 8. The items of Scale 8 consisted of all those whose cosines with the first centroid axis were .90 or above, but its own centroid vector was not computed separately.

The correlations corrected for attenuation approximate the cosines fairly well except for Scales 2 and 4. Scale 2 (Military) is much more nearly independent of the others than it was in the previous sample, and Scale 4 (Community) is somewhat more independent. The rank correlation between the 28 raw correlations and the cosines is .63; between the correlations corrected for attenuation and the cosines it is .64.

Scales 5, 6, and 7 correlate highly with Scale 8. Using the correlations corrected for attenuation, the computed multiple correlation of Scales 4, 5, and 6 with Scale 8 is 1.02; the slight inconsistencies between the reliability coefficents and the intercorrelations, and the rounding errors in the two-digit entries overshadow any difference between the true multiple correlation, corrected for attenuation, and unity. A measure of "General Morale" just as good as that given by Scale 8 can be obtained by adding half the score on Scale 7 to the scores on Scales 5 and 6.

Criterion Data

Data from Official Records

The investigators were granted access to the personnel records of the airmen designated to participate in the study. Approximately twenty items of information from the personnel folders were recorded for each airman for whom records could be located. In addition, the Medical unit provided injury data and dispensary data (frequency of visits) on the airmen for whom they had records. The following items comprised the personnel data:

1. Name
2. Rank
3. Serial Number
4. Race
5. Marital Status
6. Number of Dependents
7. Total Years of Service
8. Age
9. Education
10. Aptitude Indexes
11. Airman Proficiency Test Scores
12. Primary AFSC
13. Control AFSC
14. Duty AFSC
15. Months in Squadron
16. Citations if any
17. Airman Performance Reports
18. Proficiency Pay
19. AWOLs since 7-1-58
20. Injuries since 7-1-58
21. Dispensary visits (non-injury) since 7-1-58
22. Disciplinary Actions

The administrative officer at the Base Hospital provided for each airman information on the number of times he had been to the hospital for injuries (initial visits) during the past year, and also the number of times (initial visits) he had visited the hospital for sickness during that period.

Data from Personal Interviews

Additional data were obtained through personal interview with the immediate supervisor of each airman tested. The interview was conducted in private, and at the beginning of each interview the nature and purpose of the project was explained. The supervisor was assured that no one in the Air Force would ever see his answers to the questions in the

interview. The supervisor was then presented with a sheet of paper on which was printed a graduated scale as follows:

EXTREMELY HIGH	VERY HIGH	HIGH	LOW	VERY LOW	EXTREMELY LOW

He was asked to show the place on the scale that best described the morale of the group he supervised. Next he was asked to indicate which place on the scale best described the overall level of job performance of his group. For either rating he was permitted to place his mark *between* "High" and "Low" if he stated that his group was too near average to be properly described by either of these terms. He was then asked these same two questions separately about each airman in his group who had completed the morale questionnaire. If there were two or more of his airmen participating, he was asked to rank them with respect to overall job performance. In cases where a supervisor had only one airman participating, he was questioned until a suitable reference group was established, and then requested to rank the participating airman with respect to this reference group. The supervisors were also asked to estimate the number of times each airman had asked to be excused from duty during the previous three-month period for reasons other than his own sickness. Finally, he was asked whether or not the airman was presently doing the kind of work for which he had been trained.

Each interview took from five to twenty minutes, depending on the number of participating airmen in the supervisor's unit.

Quantification of Interview Data

The morale and performance ratings were scored on a seven-point scale:

1. Extremely high
2. Very high
3. High
4. Average (neither high nor low)
5. Low
6. Very Low
7. Extremely Low

The performance rankings were scored so that *1* represented the top man in the group. These ranks were then converted into normalized scores ($M = 500$; $\sigma = 100$) using the Fisher-Yates tables. The "time-off-from duty" score was recorded as the number of absences during the past three months as estimated from memory by the supervisor. These data were obtained as estimations by the supervisors because it was believed that the squadron records were not sufficiently uniform with respect to the recording of such absences from duty.

Validity Analyses

The "General Morale" scale (Scale 8) was validated against all available record and interview criteria. For the ones deemed more important, all eight scales were

validated. Null hypotheses (hypotheses of zero validity) were tested in most cases by the *F* test, which was used instead of the *t* test even where the latter was applicable. This was done for simplicity of computation and uniformity of reporting. Where the criterion variables were essentially continuous, scale-criterion correlations were computed.

For any given criterion item, there were in most cases some men for whom that item was not available. The number of cases varies, therefore, from one comparison to another.

In interpreting scale means and mean differences, it should be noted again that a lower scale mean indicates a *more favorable* attitude or *higher* morale.

Citations

The mean scale scores for airmen who received one or more citations during the previous year and for those who did not are as follows:

Scale	Received Citations (N = 86)	No Citations (N = 390)	F	
1	28.65	28.48	.01	
2	11.44	12.09	2.57	
3	20.71	22.82	6.80	$F_{.05} = 3.86$
4	13.44	14.27	3.06	
5	32.95	34.97	3.18	$F_{.01} = 6.69$
6	29.89	30.48	.39	
7	28.00	29.40	2.19	
8	65.21	67.53	1.67	

Scale 3 (Job Satisfaction) shows a difference significant at .01. Those who received citations like their job significantly better, but the actual mean difference is only 2.11 points or 31 per cent of the standard deviation (6.79) for the whole group of 480.

Airman Performance Reports

These are on a six-point scale with 6 the high rating. For the 295 of our 480 for whom they were available the distribution was as follows:

Rating	1	2	3	4	5	6
Frequency	0	1	26	122	129	17

Combining the first four rating categories and the last two, the General Morale means are as follows:

Rating Categories	N	M	F	
01, 02, 03, 04	146	64.16	1.46	$F_{.05} = 3.87$
05, 06	149	66.29		

The difference is not significant; and such as it is, it associates higher morale with lower performance ratings.

Proficiency Pay

The General Morale means of those who did and did not receive proficiency pay are as follows:

	N	M	F	
Received Proficiency Pay	14	66.94	.11	$F_{.05} = 3.86$
Did Not Receive	465	68.36		

The difference is insignificant.

Injuries

The correlations between the scales and number of initial visits to the Base Hospital for treatment of injuries during the previous year are as follows:

Scale:	1	2	3	4	5	6	7	8
r:	.07	.07	.07	.23	.05	.03	.05	.06

N = 403 $r_{.05} = .098$ $r_{.01} = .128$

Scale 4 (Community) has a correlation of .23, which is significantly greater than 0 beyond the .01 level, but not high enough for useful prediction.

Dispensary Visits

The correlations between the scales and the number of initial visits to the dispensary during the previous year are as follows:

Scale:	1	2	3	4	5	6	7	8
r:	.07	−.07	−.01	−.01	−.06	−.02	.03	.02

N = 403 $r_{.05} = .098$

None of the correlations are significantly different from 0 at the .05 level.

Disciplinary Actions

The mean scale scores of airmen who did and did not incur some disciplinary action during the previous year are as follows:

Scale	Disciplinary Action(s) (N = 19)	No Disciplinary Action (N = 447)	F	
1	28.21	28.50	.00	
2	11.10	11.93	1.09	
3	22.63	22.30	.00	
4	12.84	14.19	2.33	
5	32.53	34.65	.96	$F_{.05} = 3.86$
6	28.05	30.41	1.66	
7	29.10	29.06	.01	$F_{.01} = 6.69$
8	64.42	67.05	.54	

While none of the mean differences is significant at the .05 level, it is interesting to note that six of the eight differences indicate higher morale on the part of those incurring some disciplinary action.

Job Performance (Supervisors' Ratings)

The correlations between the eight scales and the supervisors' ratings of job performance are as follows:

Scale:	1	2	3	4	5	6	7	8
r:	.16	.17	.20	.04	.14	.06	.13	.09

N = 395 $r_{.05} = .098$ $r_{.01} = .128$

All of the scales except 4 (Community), 6 (Management-Communication), and 8 (General Morale) correlate significantly at the .01 level. For the three exceptions, the correlations are not significant at .05 either, though that of Scale 8 is almost significant. Despite statistical significance, the correlations are too low for useful prediction. That these correlations may be attenuated substantially by unreliability of the performance ratings is suggested by the fact that the correlation between these ratings and the performance rankings by the same supervisors is only .51.

Rated Morale (Supervisors' Ratings)

The correlations are as follows:

Scale:	1	2	3	4	5	6	7	8
r:	.19	.23	.23	.02	.24	.17	.25	.15

N = 395 $r_{.05} = .098$ $r_{.01} = .128$

All of the correlations except that with Scale 4 (Community) are significant at the .01 level or beyond, but none of them is high enough for much useful prediction. This may be due again, however, to unreliability of the ratings.

Rated Group Morale (Supervisors' Ratings)

Most of the ratings were in the above-average region, so it was necessary to combine categories to obtain substantial numbers in each. The mean General Morale scores for the combined categories were as follows:

Group Morale Rating	N	M	F	
"Extremely High" and "Very High"	52	64.46		
"High"	110	67.94	1.17	$F_{.05} = 3.04$
"Middle," "Low," "Very Low," and "Extremely Low"	47	68.98		

The trend associates high General Morale with high ratings, but it is insignificant. The correlation between the 209 individual General Morale scores and the group ratings was .12, while $r_{.05} = .136$ for N = 209.

Absences (Supervisors' Estimates)

The correlations between the eight scales and the supervisors' estimates of numbers of absences for reasons other than sickness of the men themselves are as follows:

Scale:	1	2	3	4	5	6	7	8
r:	.04	-.02	.01	.00	.11	-.01	.03	.08

N = 391 $r_{.05} = .098$

None of the correlations is significantly different from 0 at the .05 level except Scale 5, which is just barely significant. Since this is one coefficient out of eight, it too can be discounted.

Compatibility of Work with Training

The mean scale scores of the men who were and were not assigned to jobs for which they had been trained, according to supervisors' statements, are as follows:

Scale	Trained (N = 335)	Not Trained (N = 55)	F	
1	27.80	29.05	.86	
2	11.92	11.98	.01	
3	22.01	22.82	.72	
4	14.07	14.73	.14	$F_{.05} = 3.87$
5	34.66	33.62	.57	
6	30.40	29.76	.32	
7	28.93	28.50	.09	
8	66.89	65.58	.34	

None of the differences is significant.

Performance Rankings

The correlations between the eight scale scores and the normalized supervisors' rankings on performance are as follows:

Scale:	1	2	3	4	5	6	7	8
r:	-.15	-.23	-.18	-.03	-.15	-.12	-.15	-.17

N = 380 $r_{.05} = .101$ $r_{.01} = .132$

All of the correlations except the one for Scale 4 (Community) are significantly different from 0, but again they are too low for useful prediction. The fact that they are all negative is due merely

to the scoring procedure: favorable attitudes tend to go with rankings indicating high performance.

Military Rank

The mean General Morale scores for men of various ranks are as follows:

Rank	N	M
A/B and A/3C	57	71.05
A/2C	139	70.27
A/1C	138	65.77
S/Sgt	111	64.23
T/Sgt	34	60.50

N = 479 F = 5.36 $F_{.0005} = 5.10$

The tendency for men of higher rank to have higher morale is highly significant.

Size of Work Group

The mean General Morale scores for men in work groups of various sizes are as follows:

Group Size	N	M
2-3	90	66.13
4	96	66.14
5	56	63.95
6	38	71.18
7	24	62.25
8	28	75.18
9 and over	47	65.09

N = 379 F = 2.80 $F_{.05} = 2.13$

The differences as a whole are significant at the .05 level. The observed tendency for men in groups of 6 and 8 to have lower morale is not interpretable.

Race

The mean General Morale scores of White and Negro airmen are as follows:

Race	N	M	F	
White	435	67.68	8.77	$F_{.005} = 7.96$
Negro	45	60.58		

The morale of the Negroes is significantly higher than that of the Whites.

Marital Status

The mean General Morale scores by marital status are as follows:

Marital Status	N	M	F	
Married	298	66.42		
			1.15	$F_{.05} = 3.86$
Not Married	182	67.98		

This relationship is not significant.

Number of Dependents

The mean General Morale scores of airmen having various numbers of dependents are as follows:

Dependents	N	M
None	165	68.38
One	131	68.63
Two	76	66.97
Three	68	63.52
Four or more	40	62.10
N = 480	F = 2.65	$F_{.05} = 2.39$
		$F_{.025} = 2.81$

There is a just significant tendency for higher morale to go with larger numbers of dependents.

Age and Length of Service

The correlations between scale scores and age, scale scores and length of service, and scale scores and length of service with age "held constant" by partial correlation are as follows:

Scale	1	2	3	4	5	6	7	8
r_{12}	-.04	-.25	-.13	-.13	-.23	-.15	-.08	-.18
r_{23}	-.01	-.26	-.10	-.05	-.25	-.12	-.04	-.17
$r_{12.3}$	---	-.05	-.09	-.19	-.02	-.09	---	-.07

N = 480 Age = Var. 1 Morale = Var. 2 Length of Service = Var. 3

$r_{.05} = .090$ $r_{.01} = .118$ $r_{13} = .89$

The first row gives the correlations with age. All of them are significant except those for Scale 1 (Supervision) and Scale 7 (Unit). The second row gives the correlations with length of service. All of these are significant except those for Scale 1 (Supervision), Scale 4 (Community), and Scale 7 (Unit). The third row gives the correlations with length of service

when age is partialled out. The Scale 3 (Job satisfaction) and Scale 6 (Management-Communication) correlations are barely significant at the .05 level, but the Scale 4 (Community) correlation becomes significant at a level considerably beyond .01.

Education

The mean General Morale scores for airmen having various amounts of education are as follows:

Education	N	M
Grammar School or less	60	66.37
Some High School	145	66.16
High School graduate	233	68.23
At least some College	34	64.00

$N = 472$ $F = 1.08$ $F_{.05} = 3.14$

There is no significant relationship between morale and amount of education.

Aptitude Test Scores

The correlations between the General Morale scale and four aptitude indexes of the Airman Classification Battery are as follows:

	Scale								
AI	1	2	3	4	5	6	7	8	N
Technician	−.11	−.04	.00	.01	.05	.09	−.04	.03	279
Mechanical	−.02	−.04	.00	.00	.08	.13*	.03	.07	283
Clerical	−.12*	−.03	.03	.02	.02	.07	−.03	.01	282
Electronics	−.08	−.03	−.01	.03	.10	.12	−.02	.07	213

*Significant at .05 level: $r_{.05} = .118$ for $N = 282$

In general, the correlations are not significant. The two which are barely significant at the .05 level may be discounted.

Airman Proficiency Tests

As part of qualifying for change of AFSC to the next higher level in a career area, an airman takes a job knowledge test, the Airman Proficiency Test for his particular specialty. The correlations between the eight scales and scores on the Airman Proficiency Tests are as follows:

Scale:	1	2	3	4	5	6	7	8
r:	−.05	−.03	.04	−.01	−.03	−.05	−.06	−.05

$N = 343$ $r_{.05} = .107$

None of these correlations is significant at the .05 level, though all but one have the sign associated with a positive relationship between test scores and morale scores.

Assignment and Career Field

The mean General Morale scores for those whose Primary AFSC and Duty AFSC agree completely, agree partially, and do not agree are as follows:

Agreement	N	M
Complete	367	67.08
First 3 digits	70	64.20
None	43	71.02
N = 480	F = 2.63	$F_{.05}$ = 3.01

The relationship is not significant, and the trend is not consistent.

The mean General Morale scores for those whose Control AFSC and Duty AFSC agree completely, agree partially, and do not agree are as follows:

Agreement	N	M
Complete	413	66.98
First 3 digits	53	66.40
None	14	70.71
N = 480	F = .49	$F_{.05}$ = 3.01

Here again there is no significant trend.

Length of Time in Squadron

The correlations between the eight scale scores and length of time in squadron are as follows:

Scale:	1	2	3	4	5	6	7	8
r:	-.04	-.05	-.14	.08	-.04	.00	-.13	-.05

N = 480 $r_{.05}$ = .090 $r_{.01}$ = .118

The correlations with Scale 3 (Job satisfaction) and Scale 7 (Unit) are significant at the .01 level. None of the others is significant at the .05 level.

Overall Value to the Air Force

A few squadrons were asked to schedule for administration of the questionnaire their three or four *least* desirable airmen: the ones whose potentialities were least from the standpoint of the Air Force. Of the airmen selected for this sample, it was possible to match 47 with 47 other airmen who were given the questionnaire along with the regular sample. It was desired to match on age, race, rank, years of service, education, marital status and Primary AFSC. This matching was achieved with minor exceptions. The matched-pair *t* test was applied to test the significance of the mean difference on each of the eight scales. The results appear on the next page.

Scale:	1	2	3	4	5	6	7	8
t:	1.97	1.99	1.05	-.48	.90	.81	2.68	1.40

N = 47 pairs $t_{.05} = 2.01$; $t_{.01} = 2.69$ (two-sided)

For all scales except Scale 4 (Community), the "special group" had lower morale. However, only in the case of Scale 7 (Unit) was the difference significant beyond the .05 level.

Summary

1. Scale 2 (Military) has a low mean, a low standard deviation, a low reliability, and low correlations with the other scales. The airmen of this sample express such uniformly high satisfaction with the Air Force as a military organization that the variability of their scores is depressed along with the mean, and this in turn lowers the reliability and the correlations with the other scales. This is the only scale on which the members of this sample show a radically different average attitude from that shown by the members of the initial sample.

2. With the exception of Scale 2, the eight scales all have satisfactory reliabilities. The reliability of Scale 4 (Community) is only .70, but this scale has only five items. The other six scales all have reliabilities above .80.

3. The intercorrelations among Scales 1-7 inclusive are all substantially lower than their reliabilities. Scales 5, 6, and 7 correlate highly with Scale 8. Weighted 1, 1, and .5, they form together just as good a measure of General Morale as does Scale 8.

4. The reproducibilities are all in the neighborhood of .5. In Guttman's terms these scales are at best quasi-scales. All reproducibilities were computed for the 5-category scales; categories were not combined for the purpose of increasing reproducibility. The factorial homogeneity of the scales had been demonstrated in a previous study.

5. Against the criteria here employed, the validities are uniformly low. In a number of cases these validities are significantly different from zero at low probability levels, but the actual correlations and mean differences are too low for useful prediction. In the case of the supervisors' ratings, the suspicion that low validities may be due in considerable part to low criterion reliability is supported by a correlation between ratings and rankings of the same airmen by the same supervisors on the same criterion variable (production) of only .51.

Interpretations and Conclusions

Scaling

In the previous study, the primary criterion for the allocation of items to scales was factorial homogeneity. A set of items is factorially homogeneous if every item of the set has approximately the same factorial structure. The inter-item correlations should yield a Spearman hierarchy: one single general factor plus a unique factor for each item – the latter to be interpreted as error of measurement. It has been shown (Cureton, 1958) that factorial homogeneity is a necessary condition for the applicability of the generalized

Kuder-Richardson Formula 20 as a measure of the internal consistency ("instantaneous" reliability) of the scale. A secondary criterion was reasonably high communality, since the communality of an item is a close lower bound for the item's reliability. In the following brief tabulation, r_n is an estimate of the reliability of each of the eight scales obtained by taking the mean communality from the previous study (N = 1000) as an estimate of the reliability of one item and raising by the Spearman-Brown formula. KR20 is the generalized Kuder-Richardson formula 20 reliability as actually computed from the data of the present study (next-to-last row of Table 2).

Scale:	1	2	3	4	5	6	7	8
r_n:	.94	.69	.86	.72	.93	.89	.88	.95
KR20:	.92	.65	.85	.70	.91	.84	.86	.92
Diff:	.02	.04	.01	.02	.02	.05	.02	.03

In every case r_n exceeds KR20, though theory says r_n is a lower bound. The reason is probably that the previous sample was somewhat more heterogeneous, and hence more variable, than the present sample. In any event, the agreement is striking. The largest difference is .05, the mean difference is .03, and the rank correlation between the eight pairs of values is .92.

The items of a perfect Guttman scale form factorially a perfect simplex, which has $\frac{n}{2}$ common factors if n is even, and $\frac{1}{2}$ (n-1) if n is odd (DuBois, Manning, & Spies, 1959, pp. 8-23). It follows that if the items of a scale agree perfectly or almost perfectly in partially ranking the subjects, the scale *cannot* be factorially homogeneous, and the Kuder-Richardson formula 20 is *not* an experimentally unbiased estimate of its internal consistency. Internal consistency and Guttman scalability are mutually incompatible criteria for the choice of items to form a scale.

Which criterion is "better" cannot be answered until we can agree upon a criterion of "goodness." No such criterion has as yet been proposed. We may note, however, that Mahoney (1956), who did find substantial correlations between morale-scale scores and behavioral indices, used Guttman's methods in the construction of his scales. We chose internal consistency (factorial homogeneity) as the criterion, because it is clearly the appropriate criterion for the formation of scales to be used as variables in factor-analytic studies of the inter-scale correlations. This choice may well explain why our scales have good Kuder-Richardson-20 reliabilities but low coefficients of reproducibility. The use by other writers of Guttman scalability as a criterion for scale construction may also explain in part why re-factoring of their inter-scale correlations with iteration of the initial factoring procedure to produce stable communalities seems not too infrequently (e.g., in the case of Mahoney's data) to lead to a Heywood case, with one communality converging to a value greater than unity. If the number of items in each of several perfect Guttman scales exceeds the number of scales used in a factor analysis, the minimum number of factors required by the scale criteria will exceed the maximum number which can be extracted from

that number of scales. Whatever their practical values, Guttman scales are not the type of variables suitable for factor analysis. Guttman presumably recognized this in proposing the substitution of simplex, circumplex, and radex analysis for factor analysis.

Validity

The dilemma which has plagued investigators in the area of morale for some time is well illustrated in the present study. If one claims that the morale scales are all that they should be, then one must conclude that the criterion measures are either unreliable or non-relevant. If it is assumed that the criterion measures are relevant and workably reliable, then it must be concluded that the morale scales are not valid.

In the past most investigators have been content to draw one or the other of these conclusions (seldom both), and proceed as though the assumption remained tenable concerning the inherent positive relationship between expressed morale and behavioral continua representing desirable organizational behavior. The present investigators depart from this tradition. The following claims are made:

1. The measurement characteristics of the morale scales developed in the present study meet the technical requirements of good measuring devices. The items are judged to measure *expressed morale.*

2. The relevance of the criterion measures to organizational objectives is judged high.

In view of the low correlations between the morale scales and the criterion measures, the conclusion is drawn that the relationship between expressed morale and the selected measures of organizational behavior is *zero to negligible.*

Some question may then arise as to the usefulness of expressed morale measures, if such measures have no predictive efficiency for behavioral criteria. Such a question is not appropriate since in some studies, morale scales *have* had such predictive power (Mahoney, 1956). The appropriate question is, therefore, "under what condition will morale measures be related to performance measures?" Marquis, *et al.* (1951) have suggested that low morale and high productivity will occur together under the conditions of: (1) clearly-defined single goals, (2) clear-cut division of labor, (3) necessary skills well-known and possessed by group, and (4) strong pressures from outside the group itself toward conformity.

Another question that arises concerns the "validity" of morale scale measures independent of their correlations with any external criteria. This question reduces to the question of the absolute desirability of high expressed morale. While this is usually a matter of social definition, it is generally assumed that high expressed morale is "good." Certainly, the condition of high expressed job satisfaction, along with expressed feelings of personal goal fulfillment is consonant with the objectives of those who are concerned with the welfare of the individual in society. The question of whether or not the respondents in morale studies indicate their "true" or "real" feelings is meaningless, since operations do not exist for the determination of this kind of "validity."

The above results and conclusions need not be too disconcerting to those who have specialized in morale measurement. Stogdill (1959) relegates "morale" to a level of importance equal to that of the "productivity" and "integration," which three outcomes represent the dimensions of group achievement. Thus "morale" is conceived as logically independent of both integration (of the group) and productivity. All three represent group accomplishment

and represent the effects of transformation through group structure of the three basic group-input variables: performances, interactions, and expectations. From the standpoint of the present study, Stogdill's theoretical formulation is significant in that morale is assigned to the *output* end of the input-output model. This has the effect of changing radically the direction of morale research. Instead of designing studies in which morale is considered an intervening variable, or even a cause, in the cause-effect model, it is now suggested that morale studies be designed from the standpoint of attempting to delineate the *conditions* under which morale is related to other group outcomes. This in turn has the effect of putting on equal theoretical par those studies in which expressed morale measures are discovered to relate only to one another and those in which such morale measures are discovered to correlate substantially with other variables.

References

Cureton, E.E. The definition and estimation of test reliability. *Educ. psychol. Measmt,* 1958, **18**, 715-38.

Cureton, E.E. *Dimensions of airman morale.* Lackland Air Force Base, Texas: Personnel Laboratory, Wright Air Development Division, June 1960. (Technical Note WADD-TN-60-137)

Cureton, E.E., & Sargent, B.B. *Factor-analytic reanalysis of studies of job satisfaction and morale.* Lackland Air Force Base, Texas: Personnel Laboratory, Wright Air Development Division, July 1960. (Technical Note WADD-TN-60-138)

DuBois, P.H., Manning, W.H., & Spies, C.J. *Factor analysis and related techniques in the study of learning.* St. Louis; Washington University Department of Psychology, 1959. (Technical Report No. 7, Office of Naval Research, Contract No. Nonr 816(02)).

Mahoney, G.H. Unidimensional scales for the measurement of morale in an industrial situation. *Human Relations*, 1956, **9**, 3-26.

Marquis, D., *et al.* In Guetzkow, H. (Ed.) *Groups, leadership, and men.* Pittsburgh: Carnegie Press, 1951. Pp. 56-67.

Stogdill, R.M. *Individual behavior and group achievement.* New York: Oxford University Press, 1959.

Whitlock, G.H. *The current status of morale measurement.* Lackland Air Force Base, Texas: Personnel Laboratory, Wright Air Development Division, May 1960. (Technical Note WADD-TN-60-136)

Appendix

Item Numbers Identifying the Same Item in Form A and Revised Questionnaires

Form A	Revised	Form A	Revised	Form A	Revised	Form A	Revised
1	9	39	66	79	84	123	106
2	17	40	40	80	74	125	114
3	3	41	79	81	46	126	105
4	29	43	52	82	92	127	31
5	8	44	27	83	63	128	102
6	36	45	10	85	85	129	56
7	16	46	15	86	78	130	107
8	47	47	42	89	96	131	112
9	11	49	49	90	77	132	122
10	22	52	44	92	58	133	132
12	6	53	59	93	53	134	117
13	7	54	28	94	69	135	95
14	5	55	33	95	30	136	118
15	20	57	70	96	86	137	129
16	12	58	51	98	81	138	93
17	32	60	23	99	19	139	121
18	14	61	103	100	64	140	91
19	38	62	62	102	75	142	133
20	54	63	21	103	104	143	120
21	45	64	82	105	101	144	136
22	26	65	61	106	127	145	124
23	4	66	35	107	99	147	130
24	1	67	50	108	76	148	134
25	25	68	39	109	110	149	113
26	65	69	41	110	128	150	135
27	90	70	60	112	88	154	80
28	18	71	108	113	48	157	98
29	57	72	100	114	97	159	126
30	24	73	115	116	71	160	119
31	43	74	89	118	72	161	125
33	2	75	67	119	109	162	111
34	13	76	34	120	94	163	116
35	55	77	68	121	87	164	123
38	37	78	73	122	83	165	131

Div. 23/1, 28/4

UNCLASSIFIED

Wright Air Development Division. Personnel Laboratory, Lackland Air Force Base, Texas. VALIDATION OF MORALE AND ATTITUDE SCALES, by Gerald H. Whitlock & Edward E. Cureton, June 1960. 20 p. (Project 7719; Task 17130) (WADD-TR-60-76) (Contract AF 41(657)-247, University of Tennessee)
Unclassified report

This paper completes reporting of the development of scales for measuring morale among Air Force personnel. A previously developed questionnaire keyed for eight scales was given to an independent sample. When the scores were compared with criteria derived from interviews, ratings, and Air Force records, validities were uniformly low. Although none of the validities are high enough for useful prediction, the scales do measure expressed attitudes with considerable reliability.

UNCLASSIFIED

Div. 23/1, 28/4

UNCLASSIFIED

Wright Air Development Division. Personnel Laboratory, Lackland Air Force Base, Texas. VALIDATION OF MORALE AND ATTITUDE SCALES, by Gerald H. Whitlock & Edward E. Cureton, June 1960. 20 p. (Project 7719; Task 17130) (WADD-TR-60-76) (Contract AF 41(657)-247, University of Tennessee)
Unclassified report

This paper completes reporting of the development of scales for measuring morale among Air Force personnel. A previously developed questionnaire keyed for eight scales was given to an independent sample. When the scores were compared with criteria derived from interviews, ratings, and Air Force records, validities were uniformly low. Although none of the validities are high enough for useful prediction, the scales do measure expressed attitudes with considerable reliability.

UNCLASSIFIED